



# Real-Time Detection of Fraud in Credit Card using Machine Learning

Devu Kumar M S<sup>1</sup>, Narendramurthy J<sup>2</sup>, Nishant K<sup>3</sup>, Akash K<sup>4</sup>

Student, Computer Science, National Institute of Engineering, Mysore, India<sup>1-4</sup>

**Abstract:** Credit card payment has become very popular today and it is the easiest payment method along with the greatest increase in the transactions of credit card. credit card also become increasingly rampant in the recent years. Credit card fraud events are takes place recurrently and results in a huge economical loss. there are different ways of modelling credit card fraud detection problem. therefore bank and other financial organizations offers a better suitable application for detecting the fraudulent transactions with much worth and request. Fraud transactions are occurred in different ways. those various type of fraud activities a can be detect by using already applied machine learning algorithms like random forest algorithm to agree if a specific transaction is non-fraud or fraud And K-means to balance the genuine transaction in the imbalance dataset. the data used in our experiments can be taken according to the private disclosure contract.

**Keywords:** Credit Card, Fraud Detection, K-means Algorithm, Random Forest Algorithm.

## I. INTRODUCTION

In 2015, credit card transaction have generated over a \$31 trillion transactions worldwide with amount of cardholders and banks have lost over \$21 billion by fraud cases. In 2015 the credit card holders have made a 225 billion purchase transactions, a figure that purchase transaction have to be surpass 600 billion by 2025. Fraud associated with credit, debit, and prepaid cards is significant and growing issue for consumers, businesses, and the financial industry. Historically, software solutions used to combat credit card fraud by issuers closely followed progress in classification, pattern recognition and clustering. Today, most Fraud Detection Systems (FDS) continue to use increasingly sophisticated machine learning algorithms to learn and detect fraudulent patterns in real-time, as well as offline, with minimal disturbance of genuine transactions. Generally, FDS need to address several inherent challenges related to the task: extreme unbalanced of the dataset as frauds represent only a fraction of total transactions, distributions that evolve due to changing consumer behaviour and assessment challenges that come with real time data processing. For example, difficulties arise when learning from an unbalanced dataset as many machine intelligence methods are not designed to handle extremely large differences between class sizes. Also, continuous changing data require most sophisticated algorithms and robust algorithms with high tolerance for concept drift in legitimate consumer behavior. Not only can fraud detection capabilities potentially increase due to a larger scope of potential methods, the cost of development can be decreased due to reduced reliance on highly specialized niche methods, expert systems, and continued research into algorithmic methods which handle class imbalance directly.

There are most of rewards in using credit card some of them are:

- 1 Ease of Purchase.
- 2 Keep Customer credit history.
- 3 Protection of Purchases.

## II. LITERATURE REVIEW

The illegal Use of the credit card makes the very large amount of loss for the banking sector and also to the credit card customers. so the banking sector and the governments faces very big issues to find out credit card fraud so the machine learning have contain the robust and intelligent algorithm to find out credit card fraud so in this paper we have explain the how to find out the frauds using ML algorithms.

The credit card fraud can be occurred in four types:

1. Stolen or Lost card: This type of fraud can occurs when an fraudster robs the card or get an full access to the card.
2. Mail theft: This type of fraud can occurs when the fraudster get an card through mail or personal detail from bank before reaching the real card owner.
3. Counterfeit Card and Card holder not present: This type of fraud can be occurred without the knowledge of card owner. they gets the full details about the card.

Some of the Difficulties are arises while detecting the fraud in the Credit Card. some of the difficulties are: Imbalanced data, Dissimilar misclassification status, overlapping data, Deficiency of flexibility, Fraud detection is costlier,



Deficiency of typical metrics. While detecting the fraud in a credit card uses dissimilar techniques, that techniques are separated into two classes: misuse detection (fraud analysis) and anomaly detection (user behaviour analysis).

In misuse detection technique agreements with supervised methodologies job in transaction level. In this technique the fraud transactions are detected using previous or historical data. The anomaly detection technique agreements with unsupervised classification. These are based on account performance. In this technique the fraud can be detected by user behaviour because the owner and fraudster cannot behaviour same based on this the fraud can be detected.

Some of the techniques to detect the Credit Card Fraud are: Artificial Neural Network (ANN), Artificial Immune System (AIS), Genetic Algorithm (GA), Support Vector Machine (SVM), Bayesian Network, Fuzzy Logic Based System. In this paper we use K-Means and random Forest algorithm to develop Credit Card Fraud detection System.

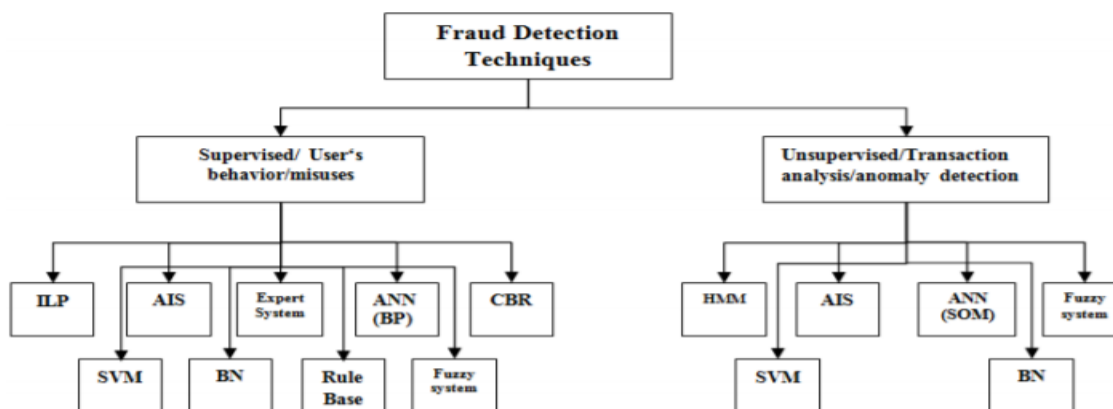


Fig.1 Classification Fraud Detection Techniques

### III. EXISTING WORK

In the current System, fraud transactions can be detected only after the transactions done. It is very difficult for the bank and other financial organizations to detect the fraud transactions which may consume more time to detect the fraudulent transactions. due to fraud activities results in a economic loss. in the earlier systems bank and other financial organizations are using may optimal classifications algorithm like bayes minimum risk. this algorithm consumes more cost with low accuracy value and as well as precision value. the major aim of the this is paper to detect the fraud activities in the real world which happens in a daily day work or business by with help of trending machine algorithms like Random forest and k-means in order to detect the whether transactions are fraud or not .by the use of this algorithms will provide greater extent of results with good accuracy and precision value.

### IV. PROPOSED WORK

In projected System, we are using two major machine learning algorithms in order to predict the fraudulent activities in the real world. two machine learning algorithms are: random forest and k-means. Random forest is one of the supervised machine learning methodology. Which is based on ensemble learning. the random forest algorithm syndicates multiple algorithm of the same type i.e multiple decision tree results in a forest trees. Hence name is called "Random Forest". this algorithm can be used for the classification and regression. it corrects the over fitting of their training dataset. subset of training dataset is sampled randomly by that to train each separate tree and then built a decision tree. Each node is splits on the feature selected form random set of complete feature dataset. it has been found to give better evaluation of simplification and to be hard to over fitting. And also, here we use K-means algorithm, K-means divides data into clusters on the basis of centroid. And there will be usually more genuine transactions(millions) then fraud transactions (a few hundred). Such data is imbalanced data and must be balanced. So genuine transactions must be balanced or labeled. No. of genuine transactions are balanced(reduced) to no. of fraud transaction. Here we are using K Means Algorithm to balance and labeled genuine transactions. There are other ways to balance data. and also, the important thing is K-means algorithm has higher scalability and efficiency.

### V. SYSTEM DESIGN AND IMPLEMENTATION

All the transaction data, ie historic data, captured from all the customers are stored in a File System. This file system could be HDFS or Amazon S3 or Azure Blob or it could be any other File System. Spark Job will import all the transaction data from the file system into the Cassandra database. This job will import fraud transactions to fraud table and non-fraud transactions to a non-fraud table.

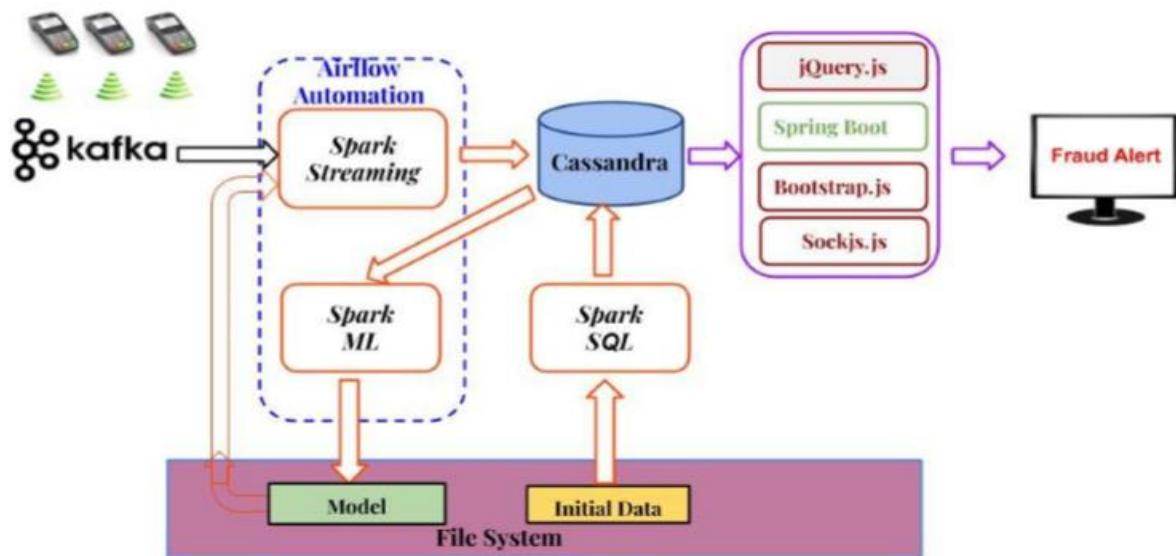


Fig.2 Architecture of the Credit Card Fraud Detection System

Spark Job will read fraud and non-fraud transactions from Cassandra. It will train on this data and it will create a model. This model will be saved to the file system. Streaming Job will load the model from the file system. It will then start consuming credit card transaction messages from Kafka. Using this model it will predict whether a transaction is a fraud or not. Finally, it will save the predictions to the Cassandra database. Fraud transactions are saved to the fraud table and the non-fraud transactions to the non-fraud table. Fraud alert dashboard will internally query Cassandra for the latest fraud transactions. If there are any fraud transactions, it will be displayed on the dashboard. Airflow Automation is used to automate Spark Streaming and Spark ML Job. Here we are automating Spark ML Job also. But in reality, Spark ML Job will be run manually by Data Science Team. It will be manually run once a week, or once a month to create a new model. The new model will be tested for efficiency. Also, it will be compared with the efficiency of the previous model. If the new model is better than the previous model then the new model will be deployed. Currently running Spark Streaming Job will be stopped and a new Streaming Job will be started to pick the new Model. Airflow automation is used to Stop and Start a new Spark Streaming Job whenever a new model is created.

## VI. CONCLUSION AND FUTURE WORK

This project intended to analyse Detection of fraud in credit card by using machine Learning algorithms. Machine learning algorithms we used in our project is Random forest algorithm and K-means algorithm. these two algorithms are very helpful while detecting the fraudulent transactions and those algorithms can be classified based on their performance of accuracy and precision. Usually, this project can be further extended to use different Classification Model must be tried and then the model efficiency must be computed using ROC( Receiver Operating Characteristics ) curve and finally models from different Algorithm must be compared using AOC( Area Under Curve ) . Once you get a good model, it must be deployed. then further it can be implemented to detect the fraud based on the location.

## REFERENCES

- [1]. Devi Meenakshi. B1, Janani. B2, Gayathri. S3, Mrs. Indira. N4: "Credit Card Fraud Detection Using Random Forest", vol.6, B1, Janani. B2, Gayathri. S3, Mrs. Indira. N4, e-ISSN: 2395-0056, p-ISSN: 2395-0072.
- [2]. Anuruddha Thennakoon<sup>1</sup>, Chee Bhagyani<sup>2</sup>, Sasitha Premadasa<sup>3</sup>, Shalitha Mihiranga<sup>4</sup>, Nuwan Kuruwitaarachchi<sup>5</sup>: "Real-time Credit Card Fraud Detection Using Machine Learning". Sri Lanka Institute of Information Technology Colombo, Sri Lanka
- [3]. <https://www.geeksforgeeks.org/ml-credit-card-frauddetection/>
- [4]. [https://www.researchgate.net/figure/Proposed-creditcard-fraud-detection-model\\_fig1\\_266746615](https://www.researchgate.net/figure/Proposed-creditcard-fraud-detection-model_fig1_266746615)